

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ**

Кафедра інформаційних систем

“ЗАТВЕРДЖУЮ”

Декан факультету інформаційних
технологій

_____ Глазунова О.Г.

“ ____ ” _____ 2019 р.

РОЗГЛЯНУТО І СХВАЛЕНО

на засіданні кафедри

інформаційних систем

Протокол № 9 від “15” квітня 2019 р.

Завідувач кафедри

_____ Швиденко М.З.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

“ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ”

галузь знань _____ 05 «Соціальні та поведінкові науки»
спеціальність _____ 051 «Економіка»
освітня програма _____ «Економічна кібернетика»
факультет _____ інформаційних технологій
розробники: _____ проф., к.е.н., Швиденко М.З., ст. викл. Стариченко Є. М.
(посада, науковий ступінь, вчене звання)

Київ – 2019 р.

1. Опис навчальної дисципліни „Інтелектуальний аналіз даних”

Галузь знань, напрям підготовки, спеціальність, освітній ступінь		
Освітній ступінь	Магістр	
Галузь знань	05 «Соціальні та поведінкові науки»	
Спеціальність	051 «Економіка»	
Освітня програма	«Економічна кібернетика»	
Характеристика навчальної дисципліни		
Вид	Вибіркова	
Загальна кількість годин	120 год.	
Кількість кредитів ECTS	4 ECTS	
Кількість змістових модулів	2	
Курсовий проект (робота) (за наявності)	-	
Форма контролю	Залік	
Показники навчальної дисципліни для денної та заочної форм навчання		
	денна форма навчання	заочна форма навчання
Рік підготовки (курс)	1	
Семестр	2	
Лекційні заняття	15 год.	
Практичні, семінарські заняття		
Лабораторні заняття	30 год.	
Самостійна робота	75 год.	
Індивідуальні завдання		
Кількість тижневих аудиторних годин для денної форми навчання	3 год.	

2. Мета та завдання навчальної дисципліни

Дисципліна "Інтелектуальний аналіз даних" належить до спеціальних дисциплін, які забезпечують формування знань та навиків магістрів економічних спеціальностей по впровадженню комп'ютерних технологій у бізнесі, основ бізнес-аналітики та її використанню у прийнятті управлінських рішень.

Мета: набуття теоретичних і практичних знань з основ інтелектуального аналізу даних.

Завдання курсу:

- оволодіння основними поняттями інтелектуального аналізу даних;
- ознайомлення з новітніми інформаційними технологіями аналітичної обробки первинної інформації;
- набуття практичних навичок по використанню інтелектуального аналізу даних у економіці.

Для вивчення дисципліни необхідні знання з основ інформатики, системного програмного забезпечення, офісного програмного забезпечення, систем управління базами даних, комп'ютерних мереж.

Засвоєння матеріалу забезпечується на лекціях, лабораторно-практичних заняттях та самостійній роботі у комп'ютерних класах, обладнаних локальними мережами, Інтернет і новітнім програмним забезпеченням. При викладанні дисципліни використовуються активні методи навчання, системний підхід, модуль-рейтингова система контролю навчання студентів.

У результаті вивчення навчальної дисципліни студент повинен:

знати: основні поняття інтелектуального аналізу даних; основні математичні моделі, що лежать в основі інтелектуального аналізу даних.

вміти: використовувати програмне забезпечення для використання процедур інтелектуального аналізу при обробці та аналізу первинної економічної інформації.

3. Програма та структура навчальної дисципліни для повного терміну денної форми навчання.

ЗМІСТОВИЙ МОДУЛЬ I. Введення в інтелектуальний аналіз даних

Тема 1. Введення в інтелектуальний аналіз даних.

Визначення інтелектуального аналізу даних (Data Mining), історія та передумови виникнення. Data Mining і витяг знань з даних (Knowledge Discovery in Databases). Місце і роль Data Mining в процесі прийняття рішень. Основні завдання Data Mining, види моделей. Data Mining і статистичний аналіз. Data Mining і сховища даних. Data Mining і OLAP. Data Mining як самостійна галузь індустрії програмного забезпечення.

Тема 2. Методи інтелектуального аналізу даних.

Класи задач (типи закономірностей), вирішуються за допомогою Data Mining: класифікація, кластеризація, регресія (прогнозування), асоціація, послідовні шаблони. Методи отримання і аналізу даних: описові - дисперсійний та регресійний аналіз; аналіз часових рядів, кластерний аналіз, еволюційні - дерева рішень, генетичні алгоритми; штучні нейронні мережі, метод «найближчих сусідів».

Тема 3. Інтелектуальний аналіз даних в СУБД MicrosoftSQLServer.

Засоби інтелектуального аналізу даних в AnalysisServices: набір стандартних алгоритмів інтелектуального аналізу даних; конструктор інтелектуального аналізу даних, призначений для створення і перегляду моделей інтелектуального аналізу даних, керування ними та побудови прогнозів; розширювану мову інтелектуального аналізу даних (DataMiningExtensionstoSQL, DMX).

Тема 4. Етапи проведення інтелектуального аналізу даних.

Етапи побудови моделі методами Data Mining: Постановка завдання, ідентифікація бізнес-проблеми, типові «вузькі місця» бізнесу. Підготовка даних, визначення джерел даних для аналізу, вибір, очищення та попередня обробка даних. Побудова моделей із зазначенням алгоритмів інтелектуального аналізу даних і їх параметри. Перевірка моделі: оцінка якості роботи створеної моделі перед початком її використання в "виробничому середовищі".

ЗМІСТОВИЙ МОДУЛЬ II. Практичні аспекти вирішення задач інтелектуального аналізу даних

Тема 5. Алгоритм дерева рішень.

Методи дерева рішень. Опис графа набору даних (вузли, гілки, листя). Основні алгоритми побудови Дерева рішень у випадку двох можливих дискретних станів результуючої змінної. «Ліси» рішень. Зростаючі «дерева» (Boosted trees) Випадкові «ліси» (Random forests).

Тема 6. Лінійні та нелінійні регресійні моделі

Статистичні методи обробки даних. Оцінювання параметрів розподілу; перевірка статистичних гіпотез; дисперсійний та регресійний аналіз; аналіз часових рядів. Класична модель ARIMA (АРПСС). Експоненційне згладжування з сезонними компонентами. Спектральне розкладання Фур'є. Поліноміальний і регресійний аналіз лагів.

Тема 7. Кластерний аналіз

Основні завдання кластерного аналізу: розробка типології або класифікації, дослідження концептуальних схем групування об'єктів, висунення гіпотез на основі дослідження даних, перевірка гіпотез. Основні етапи кластерного аналізу. Алгоритми кластеризації: м'яка і тверда кластеризація. Області застосування кластерного аналізу.

Тема 8. Нейронні мережі. Практичні аспекти вирішення задач інтелектуального аналізу даних.

Визначення та характеристика нейронних мереж. Область їх застосування. Класифікація нейромережових архітектур. Моделі нейронних мереж. Класифікація методів навчання. Особливості сучасних нейронних мереж.

Специфічні галузеві рішення Data Mining на практиці. Бізнес-завдання: в умовах жорсткої конкуренції. Застосування Data Mining при дослідженні задач: Роздрібна торгівля, Банківська справа. Телекомунікації. Страхування. Інші застосування в бізнесі

Структура навчальної дисципліни „Інтелектуальний аналіз даних”

Назви змістових модулів і тем	Кількість годин												
	Денна форма							Заочна форма					
	тижні	усього	у тому числі					усього	у тому числі				
			л	п	лаб	інд	с.р.		л	п	лаб	інд	с.р.
1	2	3	4	5	6	7	8	9	10	11	12	13	14
Змістовий модуль 1. Введення в інтелектуальний аналіз даних													
Тема 1. Введення в інтелектуальний аналіз даних	1-2	15	2		2		11						
Тема 2. Методи інтелектуального аналізу даних	3-4	15	2		4		9						
Тема 3. Інтелектуальний аналіз даних в СУБД MicrosoftSQLServer	5-6	15	2		4		9						
Тема 4. Етапи проведення інтелектуального аналізу даних	7	15	1		4		10						
Разом за змістовим модулем 1		60	7		14		39						
Змістовий модуль 2. Практичні аспекти вирішення задач інтелектуального аналізу даних													
Тема 5. Алгоритм дерева рішень	8-9	15	2		4		9						
Тема 6. Лінійні та нелінійні регресійні моделі	10-11	15	2		4		9						
Тема 7. Кластерний аналіз	12-13	15	2		4		9						
Тема 8. Нейронні мережі. Практичні аспекти вирішення задач інтелектуального аналізу даних	14-15	15	2		4		9						
Разом за змістовим модулем 2		60	8		16		36						
Усього годин		120	15		30		75						

4. Теми семінарських занять

Не передбачені навчальним планом.

5. Теми практичних занять

Не передбачені навчальним планом.

6. Теми лабораторних робіт

№ з/п	Назва теми	Кількість годин
1	Використання інструментів "AnalyzeKeyInfluencers" і "DetectCategories" надбудови інтелектуального аналізу даних для MicrosoftOffice.	4
2	Використання інструментів "FillFromExample" і "Forecast"	4
3	Використання інструментів "HighlightExceptions" і "ScenarioAnalysis"	4
4	Використання інструментів "Prediction Calculator" і "ShoppingbasketAnalysis"	4
5	Використання інструментів Data Mining Client для Excel 2007 для підготовки даних	2
6	Використання інструментів Data Mining Client для Excel 2007 для створення моделі інтелектуального аналізу даних	4
7	Аналіз точності прогнозу і використання моделі інтелектуального аналізу	4
8	Побудова моделі кластеризації, трасування і перехресна перевірка	4
Всього		30

7. Контрольні питання, комплекти тестів для визначення рівня засвоєння знань студентами.

Перелік контрольних питань

1. Data Mining як самостійна галузь індустрії програмного забезпечення.
2. Місце і роль Data Mining в процесі прийняття рішень.
3. Методи отримання і аналізу даних.
4. Поняття дерева рішень.
5. Що таке генетичні алгоритми?
6. Програмна мова інтелектуального аналізу даних.
7. Методи оцінки якості роботи створеної моделі інтелектуального аналізу.
8. Особливості дерева рішень у випадку двох можливих дискретних станів результуючої змінної.
9. Експоненційне згладжування з сезонними компонентами.
10. Спектральне розкладання Фур'є
11. Специфічні галузеві рішення Data Mining на практиці.
12. Застосування засобів інтелектуального аналізу в АПК
13. Суть системи діалогової обробки запитів
14. Сфери застосування сховищ даних.
15. В чому різниця OLAP і OLTP.
16. Класи продуктів BusinessIntelligence.
17. Інструментальні засоби бізнесу-інтелекту.
18. Алгоритм Байєса.
19. Часові ряди в економіці.
20. Поняття кластеризації.
21. Алгоритми кластеризації.
22. Особливості налаштування SQL сервера для інтелектуального аналізу.
23. Параметри точності прогнозу.
24. Методи трасування і перехресної перевірки.
25. Історія нейронних мереж
26. Типи регресійних моделей. Особливості.
27. Алгоритм дерева рішень C4.5.
28. Застосування Data mining в економіці.
29. Місце аналітика в процесі прийняття управлінських рішень.
30. Інтелектуальний аналіз даних на основі відкритого програмного забезпечення.

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ			
ОС <u>Магістр</u> Освітня програма “Економічна кібернетика”	Кафедра інформаційних систем 2019-2020 навч. рік	ЕКЗАМЕНАЦІЙНИЙ БІЛЕТ № 1 з дисципліни “Інтелектуальний аналіз даних”	Затверджую Зав. кафедри (підпис) Швиденко М.З.

1. Data Mining —

1	грубий аналіз окремих атрибутів даних (тип, довжина, спектр значень, дискретні значення і їх частота, унікальність, наявність null-значень)
2	виконує утруповання, узагальнення, пошук асоціацій, послідовностей, тобто допомагає знайти специфічні моделі у великих наборах даних
3	виявлення і видалення помилок і невідповідностей в даних з метою поліпшення їх якостей
4	вивантаження даних засобами oltp-систем в проміжні структури

2. Автоматизована ІС має забезпечувати

1	адаптацію (пристосування до прийнятої практики бізнесу та модифікації, якщо така практика змінюється)
2	підтримку професійної діяльності управлінських працівників
3	взаємодію з управлінським персоналом
4	всі варіанти вірні

3. Виберіть правильне визначення

1	Організаційне забезпечення — сукупність організаційних, методичних та технологічних документів, що регламентують процес людино-машинного оброблення інформації в АІС.
2	Інформаційне забезпечення — сукупність програм, які реалізують мету та задачі АІС і забезпечують функціонування комплексу технічних засобів системи
3	Лінгвістичне забезпечення — сукупність документів, що регламентують діяльність персоналу в АІС, взаємодію з технічними засобами і між собою в процесі розв'язування задач управління
4	Ергономічне забезпечення — сукупність методів і засобів, призначених для створення оптимальних умов високоефективної та безпомилкової діяльності людини в АІС і найшвидшого її освоєння.

4. Система діалогової обробки запитів (TPS) –

1	інформаційні додатки технології, розроблені, щоб збільшити продуктивність праці оброблювачів даних
2	комп'ютеризована система, що виконує і розраховує рутинні транзакції, необхідні для проведення бізнесу
3	сукупність програм, які забезпечують функціонування комплексу технічних засобів системи
4	інше

5 Який з програмних засобів вирішує найбільший спектр задач DataMining

1	MINITAB
2	Deductor
3	СППР "Выбор"
4	Excel

6. Сховище даних:

1	це інтерактивна, комп'ютерне заснована система, яка полегшує колективне вирішення неструктурованих та слабоструктурованих проблем групою ОГІР
2	це предметно - орієнтоване, прив'язане до часу й незмінне зібрання даних для підтримки процесу прийняття управляючих рішень

3	програмне забезпечення, що дозволяє формувати запити до даних по змісту або зразку
4	програмне забезпечення, що здійснює автоматичний пошук важливих зразків (моделей), або

7. Використання багатовимірних СУБД (OLAP) ефективне при наступних вимогах (відмітити невірне)

1	дуже великий обсяг вихідних даних для аналізу
2	набір інформаційних вимірів стабільний
3	необхідне широке використання складних вбудованих функцій для виконання кроссмірних обчислень
4	всі вірні

8. Процентне відношення кількості замовлень, що містить всі товари, які входять в правило (у замовлення можуть входити і інші товари), до загальної кількості замовлень це (написати термін)-

(у бланку відповідей впишіть вірну відповідь одним словом)
--

9. Програмні засоби підприємства, що забезпечують функції доступу і аналізу інформації, а також що забезпечують ухвалення правильних і обґрунтованих управлінських рішень -це

1	бази даних
2	BusinessIntelligence
3	OLAP
4	інше

10. BusinessIntelligence включає в себе набір програмних продуктів наступних класів (виключити зайве):

1	засоби побудови сховищ даних
2	мови програмування високого рівня
3	інструменти для виконання запитів і побудови звітів
4	системи оперативної аналітичної обробки

11. Процес виявлення в сирих даних раніше невідомих, нетривіальних, практично корисних і доступних інтерпретації знань, даних

1	статистика
2	машинне навчання
3	DataMining
4	алгоритмізація

12. Інструментальні засоби бізнесу-інтелекту

1	програмне забезпечення, що дає можливість користувачам спостерігати й використати більші обсяги складних даних
2	програмне забезпечення, що дає користувачеві можливість спостерігати дані в різних вимірах,
3	це інтерактивна, комп'ютерна система, яка полегшує колективне вирішення неструктурованих
4	програмне забезпечення, що дозволяє формувати запити до даних по змісту або зразку

13. Системи із застосуванням сучасних засобів автоматизованої обробки даних для розв'язання задач управління діяльністю підприємства

1	інформаційні системи управління підприємствами
2	інформаційно-пошукові системи
3	інформаційні системи управління технологічними процесами
4	інформаційно-довідкові системи

14. Сукупність інформаційних масивів, технічних, програмних і мовних засобів, які необхідні для збору, зберігання -

1	інформаційне забезпечення
2	інформаційна система

3	банк даних
4	інше

15. Інструментальні засоби бізнесу-інтелекту (виберіть невірний варіант)-

1	адекватність інформації стану предметної області
2	надійність функціонування
3	простота і зручність використання
4	централізації обробки інформації

16. Алгоритм класифікації, заснований на обчисленні умовної ймовірності значень прогнозованих атрибутів —

1	кластеризація
2	класифікація
3	Байєса
4	лінійна регресія

17. Ймовірність задана до досліджу

1	априорна
2	незалежна
3	несумісна
4	апостеріорна

18. Коректність використання алгоритму Байєса можлива за умов (відмітити зайве)

1	вхідні атрибути повинні бути взаємно незалежними;.
2	атрибути можуть бути тільки дискретними або дискретизованими(в процесі дискретизації безліч значень безперервного числового атрибута розбивається на інтервали і далі йде робота з номером інтервалу);
3	алгоритм вимагає меншої кількості обчислень, ніж інші алгоритми інтелектуального аналізу, що подаються Microsoft SQLServer 2008, тому він часто використовується для первинного дослідження даних. З тієї ж причини, даний алгоритм кращий для аналізу великих наборів даних з великою кількістю вхідних атрибутів.
4	всі вірні

19. Набір числових значень, зібраних в послідовні моменти часу –

1	транзакції
2	тимчасові ряди
3	часові ряди
4	неперервні ряди

20. Переваги даних в стовпчастому форматі

1	гнучкість
2	легкість машинного опрацювання
3	легше додавати та видаляти
4	менший об'єм

21. Алгоритм часових рядів Майкрософт (MicrosoftTimeSeries) включає в себе

1	Дерево авторегресії з перехресним прогнозуванням;
2	"Інтегровані ковзні середні авторегресії" (ARIMA)
3	лінійна та нелінійна регресії
4	всі перелічені

22. Який з інструментів аналітичної служби SQL Server дозволяє знизити розмірність задачі аналізу предметної області

1	кластеризація
2	групування

3	класифікація
4	агрегування

23. Алгоритм що дозволяє виявити поєднання елементів даних що часто зустрічаються і використовувати виявлені закономірності для побудови прогнозу

(у бланку відповідей впишіть вірну відповідь одним словом)

24. Який з алгоритмів кластеризації реалізує, так звану жорстку кластеризацію (варіант може належати тільки одному кластеру)

1	Алгоритм К-середніх
2	Максимізація очікувань

25. За допомогою якого інструменту можна виявити послідовності подій що часто зустрічаються

1	асоціативні правила
2	кластеризація послідовностей
3	кореляція послідовностей
4	ланцюги Маркова

26. Розташувати в порядку виконання елементи процесу навчання нейронної мережі

- Випадковим чином присвоюються значення всім вагам всіх входів в мережі.
- Для кожного навчального варіанта обчислюються виходи.
- Обчислюються помилки виходів.

27. Інструмент що дозволяє моделювати вплив, який відбувається через зміною одного з параметрів

1	Аналіз сценарію пошуку рішень
2	Аналіз сценаріїв
3	Виділення винятків
4	Аналіз можливих варіантів

28. Поставити у відповідність категорії прогнозів

А. істинний позитивний прогноз	1. клієнт, для якого прогноз показав істину, насправді зацікавлений в покупці велосипеда
Б. істинний негативний прогноз	2. клієнт, для якого прогноз показав незацікавленість у купівлі, насправді не збирається купувати велосипед
В. помилковий позитивний прогноз	3. клієнт хоче зробити покупку, хоча насправді це не так
Г. помилковий негативний прогноз	4. клієнт не хоче зробити покупку, хоча насправді він в ній зацікавлений

29. Під викидом розуміється значення даних, що є проблематичним по одній з наступних причин (відмітити невірне):

1	і значення знаходиться за межами очікуваного діапазону
2	дані представляють собою пробіл або порожній рядок;
3	значення може значно відхилитися від розподілу, якому підпорядковуються дані в моделі
4	всі вірні

30. На основі даних, що обробляються, формуються шаблони, які при використанні дозволяють віднести розглянутий приклад до однієї з можливих груп (вказати модель інтелектуального аналізу)

1	Класифікація (Classify)
2	Кластер (Cluster)
3	Пошук взаємозв'язків (Associate)
4	Оцінка (Estimate)

8. Методи навчання.

Пояснювально-ілюстративний, репродуктивний, проблемне викладання, частково-пошуковий, дослідницький.

9. Форми контролю.

Модульний контроль, поточний контроль, підсумковий контроль.

Розподіл балів, які отримують студенти. Оцінювання студента відбувається згідно положенням «Про екзамени та заліки у НУБіП України» від 27.02.2019 р. протокол № 7 з табл. 1.

Таблиця 1. Співвідношення між національними оцінками і рейтингом здобувача вищої освіти

Оцінка національна	Рейтинг здобувача вищої освіти, бали
Відмінно	90 – 100
Добре	74 – 89
Задовільно	60 – 73
Незадовільно	0 – 59

Для визначення рейтингу студента (слухача) із засвоєння дисципліни $R_{\text{дис}}$ (до 100 балів) одержаний рейтинг з атестації (до 30 балів) додається до рейтингу студента (слухача) з навчальної роботи $R_{\text{нр}}$ (до 70 балів):
 $R_{\text{дис}} = R_{\text{нр}} + R_{\text{ат}}$.

11. Методичне забезпечення

1. Інтелектуальний аналіз даних: Підручник / Черняк О.І., Захарченко П.В./ К.: Знання, 2014р. - 599 с.

12. Рекомендована література

Основні джерела:

1. Джеми Макленнен, Чжаохуэй Танг, Богдан Криват. Microsoft SQL Server 2008 :Datamining - интеллектуальный анализ данных СПб. : БХВ-Петербург, 2009
2. Кондрат Карлберг. Бизнес-анализ с помощью Excel.: Пер.с англ. – К.: Диалектика, 1997. – 448с.: ил.
3. Чубукова И. А. Data Mining: учебное пособие. — М.: Интернет-университет информационных технологий: БИНОМ: Лаборатория знаний, 2006. — 382 с

Додаткові друковані джерела:

1. А.А. Барсегян, М.С. Куприянов, И.И. Холод, М.Д.Тесс, С.И. Елизаров Анализ данных и процессов СПб.: БХВ-Петербург, 2009
2. Г.А.Титаренко. Автоматизированные информационные технологии в экономике. М.: Компьютер ЮНИТИ, 1998. – с.
3. Інформаційні системи і технології в економіці: Посібник для студентів вищих навчальних закладів/ За редакцією В.С. Пономаренка. – К.: Видавничий центр "Академія", 2002. - 544с.
4. Круглов В.В., Дли М.И. Интеллектуальные информационные системы: компьютерная поддержка систем нечеткой логики и нечеткого вывода. – М.: Физматлит, 2002.
5. Матвеев Л.А. Компьютерная поддержка решений: Учебник – СПб: "Специальная литература", 1998. – 472 с.
6. Матвеев Л.А. Компьютерная поддержка решений: Учебник – СПб: «Специальная Литература», 1998. – 472 с.
7. Плакунов М.К. Планирование на малых и средних предприятиях средствами EXCEL. – СПб.: Питер, 2004. – 156с.
8. Решение экономических задач на компьютере / Каплан А.В., Каплан В.Е., Машенко М.В., Овечкина Е.В. – М.: ДМК Пресс; СПб.: Питер, 2004.- 600с.
9. Системи підтримки прийняття рішень: Навчальний посібник/ О.І.Пушкар, В.М.Гірковатий, О.С.Євсєєв, Л.В.Потрашкова; За ред. д-ра екон. наук, проф. Пушкаря О.І.-Х.:ВД "ІНЖЕК", 2006.-304 с.

13. Інформаційні ресурси

1. Електронні навчальні курси створені в системі дистанційного навчання MOODLE , адреса – <http://it.nubip.edu.ua>)
2. Службы AnalysisServices — видеоролики по интеллектуальному анализу данных (SQL Server 2008)
3. Службы SQLServerAnalysisServices — интеллектуальный анализ данных
4. Технический справочник по алгоритму взаимосвязей (Майкрософт)
5. Технический справочник по алгоритму кластеризации (Майкрософт)
6. Технический справочник по алгоритму логистической регрессии (Майкрософт)